

Application of the combined model in short-term wind power forecasting

Liu Shijian^{a*}, Han Yongjun^b, Liu Fuchao^b

^a North China Electric Power University, Baoding, P. R. China

^b Gansu Electric Power Research Institute, Lanzhou, P. R. China

Abstract

Short-term wind power forecasting plays a major role in wind energy plant operations and the integration of wind power into traditional grid systems. It is the purpose of the present paper to provide a combined model, which is composed of the FFNN (Feed-Forward Neural Network) and LS-SVR (Least Squares Support Vector Regression) model, pertaining to the short-term wind power prediction. In this proposed approach, the FFNN and LS-SVR model can offer wind power predictions using inputs processed by PCA (Principal Component Analysis) respectively and the combined forecasting method is employed to obtain the new forecasting result. Additionally, in order to optimize the LS-SVR model, CSA (coupled simulated annealing) can assist the LS-SVR model achieve optimal performance. Our results indicate that the suggested combined model improves short-term wind power forecasts in comparison with the single models in the combined model and the traditional model.

Keywords: The combined forecasting method, wind power forecasting, LS-SVR, FFNN

1. Introduction

Wind power conceived as a sustainable energy source is gaining popularities in the electric power generation. It is estimated that by 2030, wind power could reach 2,000 GW, and supply up to 17-19% of global electricity, creating over 2 million new jobs and reducing CO₂ emissions by more than 3 billion tonnes per year [1]. Consequently, wind energy is becoming an important component in the supply mix to meet the growing demand for electric energy. However, owing to the uncertainty and volatility of wind power, large-scale integration of wind power in the wind farms will bring serious impact to the power system [2].

An accurate and reliable Wind Power Forecasting (WPF) model [3] is therefore essential for planning of economic load dispatch and grid management. It is in recent years that various methods have been proposed in the literature, such as physical methods, time series models and soft computing approaches, and applied to the domain of wind power forecasting. The physical methods make use of abundant weather data to forecast wind power. The Numerical Weather Prediction [4], [5] model is perceived as a typical physical approach to address the prediction problem.

Time series models are widely used in the forecasting field and have also been proposed for short-term wind power forecasting. The models are established using historical data to tune the model parameters and by examining whether the fitting residuals possess the characteristics of a random walk process. Typical examples of time series models include the ARMA (autoregressive moving average) [6], the ARIMA (autoregressive integrated moving average) [7], exponential smoothing techniques [8] and grey predictors [9], [10].

Soft computing methods are extensively utilized by scholars to forecast wind power because such methods provide suitable performance capabilities, especially in tackling nonlinear problems. ANNs

* Manuscript received April 15, 2016; revised July 30, 2016.

Corresponding author. Tel.: 15033121861; *E-mail address*: liushijian1992@163.com.

doi: 10.12720/sgce.5.3.144-152

(Artificial neural networks) are the most popular approaches among soft computing methods, which are based upon empirical risk minimization and asymptotic theories. Among artificial neural networks, the BPNN (Back-Propagation Neural Network) [11]-[13], the recurrent neural networks [14], the Elman neural network [15], the GRNN (General Regression Neural Network) [16], the FFNN (Feed-Forward Neural Network) [17] and MLP (Multi-Layer Perceptron network) [18] are among the most extensively used approaches for the prediction of wind power. In the same line with artificial neural networks, SVM (Support Vector Machine) [19]-[24] is another type of soft computing methods, which is based on a statistical learning theory and the structural risk minimization principle.

In this paper, a novel ensemble method consisting of PCA (Principal Component Analysis), FFNN, LS-SVR (Least Squares Support Vector Regression), CSA (coupled simulated annealing) and the weighted average of the combined forecasting method is proposed to forecast the wind power. The main contributions of this paper are as follows.

- PCA is not only utilized to reduce the dimension of the input variables but also to account for the variability of the input variables as far as possible. Before forecasting, PCA is used on the raw data to select the informative input variables efficiently.
- CSA is introduced to as an optimization method for LS-SVR, which can seek the optimal parameters in LS-SVR. These parameters are important for performance of LS-SVR, and CSA is developed to assist this ensemble method enhance the accuracy of prediction.
- Two kinds of forecasting results can be obtained by means of FFNN and LS-SVR. Furthermore, these different results can be combined reasonably to get the new result.

This paper is organized as follows. We review the relevant theories and describes the proposed method in details in Section 2. The numerical results and discussions are provided in Section 3. Finally, conclusions are drawn in Section 4.

2. The Proposed Method

2.1. Overview

Wind power forecasting is a challenging task because it is fluctuant and unable to find the general law. The wind power data and wind speed data in history, which are strongly related to the forecasting object, basically act as the input variables. However, the excess historical data may lead to information abundance and make it hard to recognize the key patterns beneath data. A key issue for the success of forecasting method is the suitable process of input variables. In the proposed method, PCA is presented to select the informative input variables efficiently.

After the key features are selected, a three-layer feed forward neural network is constructed. This kind of network contains an input layer, an output layer and a hidden layer. Network with this structure is proved to have the ability to approximate any function [25]. The adaptive and data driven approach is suitable for complex system such as wind power prediction. In order to enhance the robustness of the forecasting model, a 10-fold cross-validation is applied in this paper. SVM is similar with neural network, but it can make the structural risk minimum. The generalization ability of learning machine is greatly improved owing to the establishment of the SVM model. When it comes to the scarce training samples or the weak robustness of training system, SVM still obtains study effect of the small error [26]. Therefore, LS-SVR is introduced in this forecasting model. As for the parameters in LS-SVR, CSA will seek their optimization.

So far, FFNN and LS-SVR are presented and they have advantage and weakness respectively. In order to combine these forecasting models, the weighted average of the combined forecasting method is employed to achieve quality results.

2.2. Principal component analysis

Ever since the inception of Principal Component Analysis by Pearson [27], it gained popularity as a dimensionality reduction technique. PCA presumes that the observations can be projected on a new set of

axes, removing data redundancy and system noise in the process. Principal components are linear combinations of the original variables. The principal components are the directions that comprise the new coordinate axes. It is supposed that there are n observations and p different attributes in every observation.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = (X_1, X_2, \cdots, X_p) \quad (1)$$

The first principal component can then be formulated as a linear combination of the original variables explaining a maximum amount of variation, as follows:

$$F_1 = A_1 X = a_{11} X_1 + a_{21} X_2 + \cdots + a_{p1} X_p \quad (2)$$

It can be shown that the principal components of A are the eigenvectors of the covariance matrix of X . Naturally, there would be little use for PCA if all P components were retained. Hence, a criterion needs to be put forward that selects a number of principal components that explain a reasonable amount of variation. Scree plot can describe the eigenvectors and their sequence. In this plot, there is an inflection point that is smaller obviously than previous points. Therefore, these points before the inflection point are selected as principal components' eigenvalue.

2.3. Feed-forward neural network

First introduced by McCulloch and Pitts [28], artificial neural network has gained increasingly popularity due to its inherent ability to capture nonlinearity and complexity. ANN is a kind of frameworks of computing flexibility which can be used to model a wide range of nonlinear problems. It is a universal approximator with high degree of accuracy and no prior assumption of the model form. The basic idea of ANNs is to update the coefficients in neural network until a desired forecast error is reached. Among the various kinds of ANNs, the three-layer feedforward neural network is most widely used model for forecasting.

Before the forecasting procedure begins, network structure should be determined at first. As the input variable is determined by the feature selection procedure and the output variable is wind power, only the number of neurons in hidden layer needs to be specified. More neurons can help to reduce training error while leading to an over-fitting problem. There is no theory for neurons selection and trial and error approach is used in the proposed method. Neuron number varies from 2 to 10 until the optimal result is obtained based on out-of-sample prediction error. We constructed this neural network incorporated with Levenberg-Marquardt algorithm for training which can give a high accuracy with less time cost when compared to the classic Backward-propagation algorithm.

2.4. Least squares support vector regression

The modified version of SVM called LS-SVR [29] resulted in a set of linear equations instead of a quadratic programming problem. The brief reviews of the LS-SVR algorithm for regression problems are show as follows. Given a training set $\{x_i, y_i\}$, the regression formula can be constructed as follows,

$$y = W^T \phi(x) + b \quad (3)$$

where W denotes the weight vector and b is the bias term. And $\phi(x)$ is the nonlinear mapping function that transfers the input to a higher dimensional feature space. The weight vector W of the regression can be calculated by optimizing the following cost function containing a penalized regression error:

$$C = \frac{1}{2} W^T W + \frac{1}{2} C \sum_{i=1}^N e_i^2 \quad (4)$$

Subject to

$$y_i = W^T \phi(x_i) + b + e_i \quad (5)$$

The error variable at time t and g is a regulation constant. Based on the subjections, the Lagrange function is constructed as follows:

$$L(W, b, e, \alpha) = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i (W^T \phi(x_i) + b + e_i - y_i) \quad (6)$$

2.5. Coupled simulated annealing

The CSA (Coupled Simulated Annealing) algorithm [30] is an advanced version of the SA (Simulated Annealing) algorithm and is characterized by a series of parallel SA processes. SA suffers from the problems of the premature and the sensitivity of the initialization parameters. The CSA is designed to upgrade the quality of solutions at the minimal cost of decreasing the convergence speed. In each optimization process of CSA, each single current state is performed separately and behaves as a single classical SA process. The main difference between SA and CSA lies in the acceptance probabilities. The SA process adopts the importance sampling technique to choose sample states of a particle system model to efficiently estimate physical quantities that are related to the system. In terms of the master equation of a thermodynamic system, this principle states that

$$\frac{P(x \rightarrow y)}{P(y \rightarrow x)} = \frac{\exp(-E(y)/T)}{\exp(-E(x)/T)} \quad (7)$$

where $P(x \rightarrow y)$ denotes the probability of transitioning from the current state x to a candidate state y , T is the setting temperature, $E(x)$ and $E(y)$ denote the energy of the current state x and current state y , respectively. The transfer probability can be represented as the product of a generation probability and an acceptance probability, i.e. $P(x \rightarrow y) = G(x \rightarrow y)A(x \rightarrow y)$. If all candidate states take equal probabilities, i.e., the generation probability $G=1/n$ with n denoting the number of possible states, then formula (7) can be simplified as follows:

$$\frac{A(x \rightarrow y)}{A(y \rightarrow x)} = \frac{\exp(-E(y)/T)}{\exp(-E(x)/T)} \quad (8)$$

The most common functions for acceptance probability are the Metropolis rule

$$A(x \rightarrow y) = \exp\left(\frac{E(x) - E(y)}{T_k^{ac}}\right) \quad (9)$$

And the rule

$$A(x \rightarrow y) = \frac{1}{1 + \exp\left(\frac{E(x) - E(y)}{T_k^{ac}}\right)} \quad (10)$$

SA only considers the current solution for the acceptance decision of the probing state, whereas CSA

considers several current states and accepts a probing state based not only on the corresponding current state but also on the coupling term. The coupling can not only interchange information and produce cooperative contributing to the decision of whether uphill moves are accepted but also provide information for the entire optimization process toward the globally optimal solution.

2.6. The weighted average of the combined forecasting method

It is supposed that there are m forecasting models which will predict the same object. And then the combined forecasting model which comprises every single model is

$$f_t = \sum_{i=1}^m k_i f_{it} \quad (11)$$

where f_t denotes the prediction value of the combined forecasting model at t , f_{it} is the prediction value of the i th forecasting model at t , k_i is the weight of the i th forecasting model, and it is satisfied

$$\sum_{i=1}^m k_i = 1, k_i \geq 0 \quad (12)$$

In the weighted average of the combined forecasting method, what is the important thing is seeking the weight of every forecasting model. Here is the procedure of this method.

First, $n \times m$ matrix which means n forecasting units and m forecasting methods is constructed. The element x_{ij} of this matrix represents evaluation result of the j th forecasting method in the i th forecasting unit. After the normalization, x_{ij} becomes z_{ij} .

What is more, the numerical sequence $C_0 = (C_{01}, C_{02}, \dots, C_{0n})$ that can judge which is superior is constructed.

$$C_{0i} = \frac{\sum_{j=1}^m z_{ij}}{m} \quad (13)$$

Finally, Pearson correlation coefficient of the normalization sequence and the numerical sequence is calculated and the weight of every forecasting model is obtained.

$$k_i = \frac{\rho_{i0}^t}{\sum_{j=1}^m \rho_{j0}^t} \quad (14)$$

Larger t is, more the gap in the weights of every forecasting method is. However, t is unable to be much considerable, which may cause the bias.

3. Case Study

In this section the efficacy of the proposed approach is demonstrated with a case study and four subsections: collection of data, evaluation criteria of forecasting performance, simulation and comparison and discussion, which are presented sequentially.

3.1. Collection of data

In this paper, wind data from a wind farm located in China are used to demonstrate the effectiveness

and reliability of the proposed combined forecasting approach. Because there is a strong relationship between wind power data and wind speed data, this study samples wind power data and wind speed data in the wind farm for the time period in January, 2014. The data of the previous hour is informative explanatory variable for wind power forecasting. Consequently, it is perceived as an input for FFNN and LSSVR. Every sample includes the current wind power and the wind power and speed data of the previous hour. In this simulation, there are 300 samples for training and 10 samples for testing.

3.2. Evaluation indices for forecasting performance

To evaluate the performance of the proposed approach, three statistical indices are employed to measure the forecasting accuracy. They are the MAE (mean absolute error), RMSE (root mean square error) and MAPE (mean absolute percent error). These indices are defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |t_i - p_i| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N |t_i - p_i|^2} \quad (16)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{t_i - p_i}{t_i} \right| \quad (17)$$

where t_i is the observed value for the time period t and p_i is the predicted value for the corresponding period. The MAE reveals how similar the predicted values are to the observed values, whereas the RMSE measures the overall deviation between the predicted values and the observed values. The MAPE is a unit-free measure of accuracy for the predicted wind series and is sensitive to small changes in the data.

3.3. Data preprocessing

The excess number of input, which is 10 in total, may cause the more time in training for FFNN and LSSVR. Moreover, the input variables just contain the wind power and wind speed data. Consequently, there is redundancy information among the input variables. It is reasonable to abstract the effective information and reduce the number of input. PCA is an appropriate approach to deal with this problem. After PCA, the number of input is reduced to 6 and the most information are reserved. Furthermore, the available inputs of the combined model were linearly normalized in the range $[0,1]$ to overcome the saturation phenomenon.

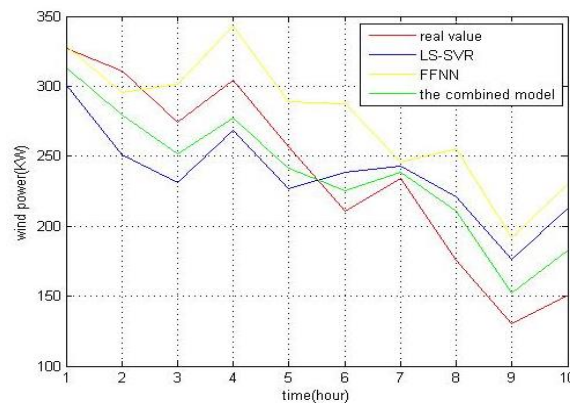


Fig. 1. The hourly forecasting results by the FFNN, LS-SVR and the combined forecasting method.

3.4. Forecasting

After the data preprocessing, the reconstructed data was processed into the input of the combined model. Therefore, 300 samples were selected as the train set for FFNN and LS-SVR. As for the parameters in LS-SVR, it is difficult to figure out the specific value. However, CSA is aimed to seek the optimization value. Hence, these two forecasting models are established after training and the forecasting results of the two forecasting models were achieved. Finally, the weighted average of the combined forecasting method is employed to combine FFNN and LS-SVR for receiving the combined forecasting results. Fig. 1. presents the hourly forecasting results by the FFNN, LS-SVR and the combined forecasting method.

3.5. Comparisons and discussion

To facilitate the analysis and discussions of the combined model, four models for short-term wind power forecasting were employed for comparison with the proposed model and the assessment of the prediction performance in this subsection. The established models were the AR (auto-regression model), the FFNN model, the LS-SVR model and the CSA-LSSVR model. In these four models, the AR model was the traditional forecasting model and others were the sub-models of the combined model.

Table 1. The evaluation results obtained from different models

models	MAE	RMSE	MAPE
AR	45.69	53.36	23.43
FFNN	42.41	50.92	16.63
LS-SVR	38.98	43.67	15.23
CSA-LSSVR	38.45	41.58	14.87
The combined model	21.75	23.57	9.72

Table 1 shows the evaluation results obtained from the AR, FFNN, LS-SVR, CSA-LSSVR and the combined models with respect to the hourly wind power prediction. It can be readily seen from Table 1 that the proposed hybrid approach outperforms the other models in terms of the three forecasting evaluation indices (MAE, RMSE and MAPE).

Contrasted with the traditional forecasting approach, the less error can be seen in the soft computing methods. Furthermore, the CSA-LSSVR model performed better than the LS-SVR model. Additionally, the FFNN model and the CSA-LSSVR model can be combined using the weighted average of the combined forecasting method and the better forecasting results will emerge.

More detailed analyses were performed. The proposed combined approach performed better than the other models in short-term wind power forecasting. For example, the developed method outperforms the other models with lower RMSE value of 23.57KW for one-hour ahead forecasting in contrast to 53.36KW, 50.92KW, 43.67KW and 41.58KW for the AR, FFNN, LS-SVR and CSA-LSSVR models, respectively. A low MAPE value of 9.72% was obtained by the proposed model, while the AR, FFNN, LS-SVR and CSA-LSSVR methods resulted in higher MAPE values (23.43%, 16.63%, 15.23% and 14.87%), respectively. The comparison of the predictions shows that the integration of the proposed data preprocessor and the forecasting engine is a good choice for short-term wind speed prediction.

4. Conclusions

With the rapid growth of wind energy, accurate and reliable methods and techniques for short-term wind power forecasting are urgently needed. Owing to the effect of various environmental factors, wind power data present high fluctuations, autocorrelation and stochastic volatility, making it difficult to forecast wind power using a single model. This paper proposed a hybrid forecasting approach based on the concept of combined prediction, which integrates the models of FFNN and CSA-LSSVR. For the prediction, CSA-LSSVR has the better prediction effort than LS-SVR in the light of the case study. And it is proved that the CSA algorithm can optimize the LS-SVR model's parameters effectively. Although

the greater error can be obtained from the traditional forecasting model in comparison with FFNN and CSA-LSSVR, the level of the error is still high. In order to gain the better forecasting results, it is feasible to combine these two models. Furthermore, the combined model overcomes the shortages of FFNN and CSA-LSSVR. Compared with these two models, results of this study indicate that the prediction accuracy is greatly improved by utilizing the combined model. Therefore, the combined model is better for short-term wind power forecasting. The effectiveness of the proposed model was demonstrated with real mean hour wind power data.

Acknowledgements

This research was supported by Gansu Electric Power Research Institute.

References

- [1] Global Wind Energy Outlook 2014, Global Wind Energy Council, Brussels, Belgium; 2014.
- [2] Zhang ZS, Sun YZ, Cheng L. Potential of trading wind power as regulation services in the California short-term electricity market. *Energy Policy*, 2013; 59:885-897.
- [3] Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. *Renew Energy*, 2012; 37:1-8.
- [4] Guannan Q., et al. Short-term wind power forecasting based on numerical weather prediction adjustment. In: *Proc. of 11th International Conference on Industrial Informatics*, 2013:29-31.
- [5] Xu Q, et al. A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. *IEEE Transactions on Sustainable Energy*, 2015; 6(4):1283-1291.
- [6] Cao Y, et al. Wind power ultra-short-term forecasting method combined with pattern-matching and ARMA-model. Presented at: 2013 IEEE GCPT.
- [7] Shi J, et al. Short-term wind power generation forecasting: Direct versus indirect ARIMA-based approaches. *International Journal of Green Energy*, 2013; 8(1):100-112.
- [8] Cadenas E, et al. Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method. *Renewable Energy*, 2010; 35(5):925-930.
- [9] El-Fouly THM, et al. Grey predictor for wind energy conversion systems output power prediction. *IEEE Transactions on Power Systems*, 2006; 21(3):1450-1452.
- [10] Shi N, et al. Wind speed forecasting based on grey predictor and genetic neural network models. In: *Proc. of 2nd International Conference on Measurement, Information and Control*, 2013:16-18.
- [11] Jiandong M, et al. Wind power forecasting based on the bp neural network. *Applied Mechanics and Materials*, 2013; 341-342(2):1303-1307.
- [12] Lingling L, et al. Wind power forecasting based on time series and neural network. In: *Proc. of Second International Symposium Computer Science and Computational Technology*, 2009:26-28.
- [13] Wen-Yeou C. Application of back propagation neural network for wind power generation forecasting. *International Journal of Digital Content Technology and its Applications*, 2013; 7(4):502-509.
- [14] Qing C, et al. Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research*, 2012; 221(1):148-154.
- [15] Xiaoyu Z, et al. Short-Term forecasting of wind power generation based on the similar day and elman neural network. In: *Proc. of IEEE Symposium Series on Computational Intelligence*, 2015:7-10.
- [16] Hong CM, Cheng FS, Chen CH. Optimal control for variable-speed wind generation systems using general regression neural network. *Electr Power Energy Syst.*, 2014; 60:14-23.
- [17] Bhaskar K, Singh SN. AWNN-assisted wind power forecasting using feed-forward neural network. *IEEE Transactions on Sustainable Energy*, 2012; 3(2):306-315.
- [18] Wei-Chang Y, et al. Forecasting wind power in the Mai Liao Wind Farm based on the multi-layer perceptron artificial neural network model with improved simplified swarm optimization. *International Journal of Electrical Power & Energy Systems*, 2014; 55:741-748.
- [19] Wendan Z, et al. A hybrid EMD-SVM based short-term wind power forecasting model. In: *Proc. of IEEE PES Asia-Pacific Power and Energy Engineering Conference*, 2015:15-18.
- [20] Wen J, et al. Short-term wind power forecasting based on lifting wavelet, SVM and error forecasting. Presented at: 2012 ICEE.
- [21] Qiang W, et al. A multivariate wind power forecasting model based on LS-SVM. In: *Proc. of Fifth International Joint Conference on Computational Sciences and Optimization*, 2012:23-26.

- [22] Yang X, *et al.* Short-term wind power forecasting based on SVM with backstepping wind speed of power curve. Presented at: 2012 ICIDMP.
- [23] Jianwu Z, Wei Q. Support vector machine-based short-term wind power forecasting. In: *Proc. of IEEE/PES Power Systems Conference and Exposition*, 2011:20-23.
- [24] Li X, *et al.* Short-term wind power forecasting based on least-square support vector machine (LSSVM). Presented at: 2013 REET.
- [25] Wang SYL, Kin KL. A neural-network-based nonlinear metamodeling approach to financial time series forecasting. *Applied Soft Computing*, 2009; 9(2):563–574.
- [26] Jie Y. Ensemble SVM regression based multi-view face detection system. In: *Proc. of Machine Learning for Signal Processing*, 2007:16-18.
- [27] Pearson LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901; 2(11):559–572.
- [28] Warren SM, Walter P. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 1943; 5(4):115–133.
- [29] Qiao M, *et al.* Time-Series gas prediction model using LS-SVR within a Bayesian framework. *Mining Science and Technology*, 2011; 21(1):153-157.
- [30] Yalan Z, Chen L. Coupled simulated annealing with differential evolution. Presented at: 2011 ICACI.